European Online Journal of Natural and Social Sciences 2015;
Vol.4, No.3 pp. 518-526
ISSN 1805-3602

www.european-science.com

MF-PCA continuum fitting of SDSS-DR7 at high redshifts in the Lyα forest

Z. Kiamehr* and A. Sardarzaei
Chabahar Maritime University, Faculty of Marine Sciences, Department of Basic Sciences
E-mail: ghaffarzadeh.mail@gmail.com

Received for publication: 10 March 2015.
Accepted for publication: 05 July 2015.

Abstract

This paper is mainly concerned with mean flux regulated - principal component analysis (MF-PCA) continuum estimation technique. In this technique PCA fitting is carried out red ward of quasar Lyα line in order to provide a prediction for the shape of the Lyα forest continuum. The purpose of this paper is also to predict the continuum of bright QSO in the region by wavelengths longer than Lyα emission line using the statistical method of MF-PCA. The method applied to a sample of 892 high redshift with brightly luminous QSO; ($M_B \leq -28$, S/N > 10 and $2.7 \leq z \leq 4.7$) published in SDSS-DR7. The slope and amplitude of the continua were predicted by applying the estimated technique. We found that MF-PCA method reduces the errors to 3% RMS in S/N > 10 spectra with continuum of manual method.

Keywords: intergalactic medium, absorption lines, emission lines, techniques of data analysis.

Introduction

In the past two decades, the Lyα forest absorption lines observed in the region of quasars with high redshift ($z \geq 2$) as an important exploration of large scale structure and the IGM (Whitney et al., 1983 and Schneider et al., 1991). While almost all aspects of the Lyα forest are dependent on the determination of the continuum (Francis et al., 1992). Therefore, the accurate prediction of the continuum quasar is vital in order to take new Lyα forest data sets. For example, the Lyα flux probability distribution function used to constrain the temperature - density relation (TDR) of the IGM and is very sensitive to continuum errors. Lee et al., (2011) has showed that 2% systematic errors in the continuum estimation can double the errors in the TDR. In other words, measurements of the one – dimensional flux power can be converted by the Fourier power introduced by quasar emission lines in the Lyα forest region. For example, the counted continuum variance in the quasar has limited the measurements of one – dimensional the Lyα forest at redshift $z = 2.5$. The exact quasar continuum fitting is a fundamental problem at redshifts $z \geq 2$. Where the Lyα forest becomes accessible to ground based optical telescopes the high absorber line density makes it challenging to observer the quasar continuum and usually estimated by same form of spline fitting to the observed transmission peaks reach the quasar continuum at a given redshifts (Press & Rybicki 1993, Songaila & Cowie 1996 and Rauch et al., 1998). These methods cannot in general be applied to large data sets such as the SDSS, as the modest resolution and low signal to noise ratio make it impossible to fit the Lyα forest and even then steps need to be taken to account for the degradation of transmission peaks from the lower resolution. This direct fitting technique usually require significant human intervention and more often is very time consuming (preventing their application to the ~ 10^4 Lyα forest spectrum lines in SDSS) (Petitjean & Ledoux 1999 and Nicolson 1999).

The data for the Lyα forest noisy is usually the best approach and using the spectrum that are less absorption lines and the redshift (the Lyα emission line on the right) are continuous the Lyα forest (usually in the range of wavelengths from 1040 to 1180Å known) includes weak emission lines such as FeII with wavelength 1071Å and FeII/FeIII with wavelength 1123Å although the exact identifications varies from author to author, these emission lines can cause deviations of up to 10% from a flat continuum. Therefore the use of an average continuum shape would not account for variation of up to 10% within individual quasars (due to the presence of emission lines) is not enough. However, there are many variations in the spectra predicted by the equivalent width of the emission lines is poor. One possible way to improve quasar continuum fits is principal component analysis (PCA). A sample of 50 quasars with low redshift observed in the UV by the Hubble Space Telescope (HST) in which the $\lambda_{Rest} < 1216$Å continuum can be clearly shown. They concluded that while PCA fits to the red side ($\lambda_{Rest} = 1216 - 1600$Å) which gave a good estimation of the Lyα continuum shape, the overall amplitude had about 10% errors (Suzuki et al., 2003 and 2005).

Paris et al., (2011) carried out a similar analysis with high signal to noise ratio ($S/N \geq 10$) quasar samples of the SDSS. They found a better prediction accuracy of 5% due to larger spectral $\lambda_{Rest} = 1025 - 2000$Å as approach to $\lambda_{Rest} = 1025 - 1600$Å in the earlier work (Paris et al., 2011). Among the disadvantages of this method is that the standard PCA method does not include the pixel noise, while the majority of the quasars observed in SDSS have low signal to noise ratio ($S/N < 10$). Francis et al., (1992) argued that PCA errors in fitting scale directly with the noise level, which shows that, for example, not better than 20% continuum accuracy in an $S/N = 5$ to be expected (Nusser & Haehnelt 2000, Becker et al., 2001, Hewett et al., 2001 and Strauss 2002). Due to all the reasons stated above, the more precise continuum fitting techniques are desperately needed to take the advantage of the Lyα forest data from SDSS and spectroscopic survey in the future (Efstathiou et al., 2002, Bernardi et al., 2003, Seljak & Cen 2003 and White et al., 2003). In summary, the least square fitting of PCA and the red part of the absorbed quasar spectrum ($\lambda_{Rest} > 1216$Å) was used to predict the continua of the mean flux expected $\langle F \rangle(z)$, to constrain the amplitude of the fitted continuum. Tytler et al., (2004) showed that the dispersion of the expected flux $\langle F \rangle$ with $\Delta z = 0.1$ part of the Lyα forest at $z = 2$ is $\sigma_F(\Delta z = 0.1) \approx 4\%$, so the Lyα forest expected range of the continuum amplitude to be ~ 2% error (Aghaee et al., 2006 and Lee et al., 2012).

Materials and Methods

The MF-PCA method which developed in this article is optimized towards large sets of noisy data spectra. This method used the SDSS data which consists of about 10^4 spectra of moderate resolution ($R \approx 2000$) and modest signal to noise ratio was also used (Paris et al., 2011). An overview of the Lyα forest data sample was used to fit the data set. There are 892 out of 105000 QSOs, published in the SDSS-DR7 which have an absolute magnitude less than -28 in the blue filter, S/N > 10 and a redshift range 2.7≤ z ≤ 4.7 have been selected. Since the extreme blue part of the spectrum (near $\lambda_{Obs} \approx 3800$Å) of the SDSS spectra are known to suffer from spectrophotometric problems of $\lambda_{Obs} = 3840$Å as the lowest limit wavelength was used. This sets a minimum quasar redshift of $z_{QSO} = 2.7$ and the parts of the spectra that lie below $\lambda_{Obs} = 3840$Å have been ignored (Kirkman et al., 2005 and Kim et al., 2007).

Using quasars with high resolution (R~2000) the focused spectra in SDSS are calibrated. The quasars that studied in here, with a full range of the Lyα forest (the wavelengths range from 1020 - 1600Å) and high signal to noise ratio were selected (Lee et al., 2012).

Fitting continua by principal component analysis (PCA) method

The basic concept of principal component analysis of normalized quasar spectra $f_i(\lambda)$ are well reconstructed by the spectrum of $r_{i,m}(\lambda)$ which is shown as below:

$$f_i(\lambda) \cong r_{i,m}(\lambda) = \mu(\lambda) + \sum_{j=1}^{m} c_{ij}\varepsilon_j(\lambda) \quad (1)$$

where $\mu(\lambda)$ is the mean quasar spectra, $\varepsilon_j(\lambda)$ is the j-th principal component and c_{ij} is coefficient of the weights of the quasar. To obtain an estimate for the continua quasar with redshift 2.7≤ z ≤ 4.7, S/N > 10 and $M_B \leq -28$ was used. Since in these spectrum low absorption existed that one can designate sufficient clarity of the Lyα forest. The quasars studies were used only in the wavelength range of the spectrum that were involved 1020-1600Å, 1020Å< λ< 1215Å blue portion of the spectrum and 1216Å< λ <1600Å red part of the spectrum (Suzuki et al., 2003 and 2005).

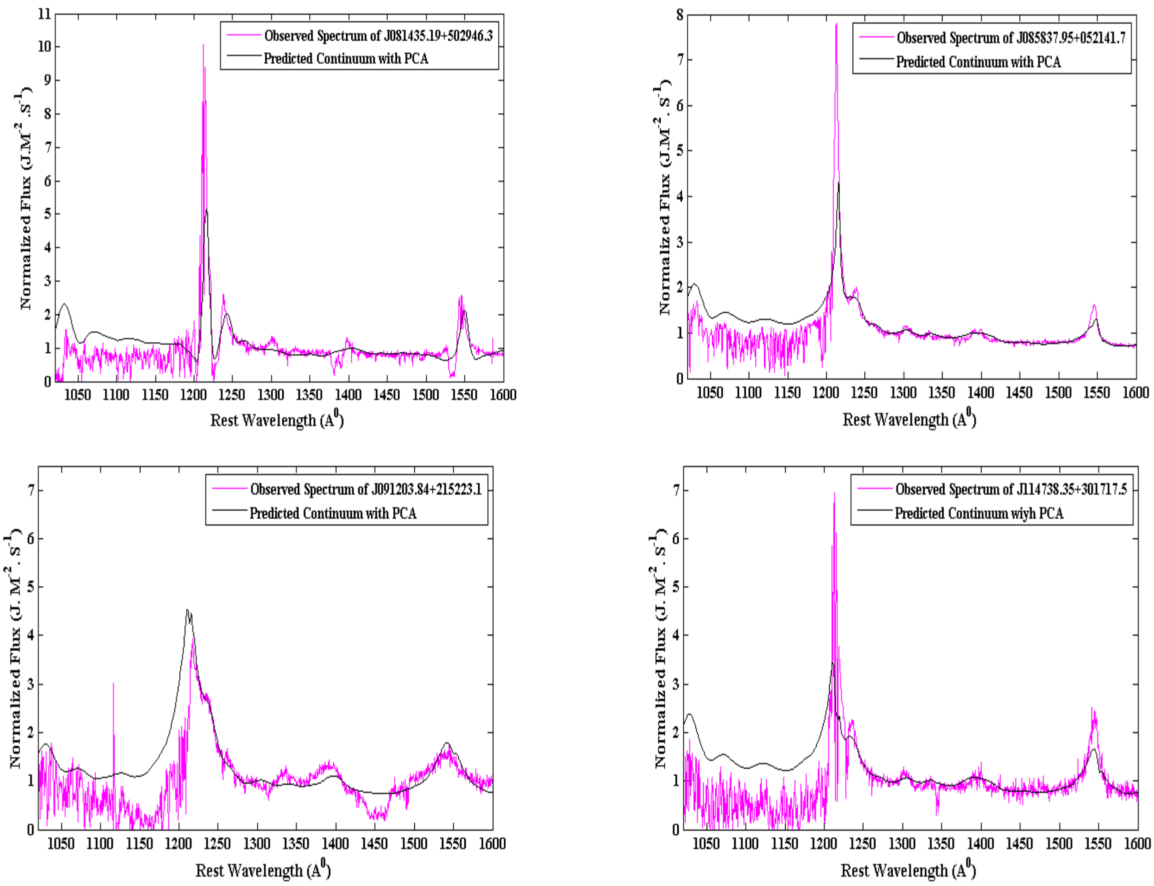

Figure 1a. The bad continua of 4 QSOs derived from applying the PCA method described in the text.

The spectra of all quasars were fitted by smooth curves, in the way that the absorption lines in both the blue and red spectrum was removed and the noise effects of input was also reduced. The transfer of redshift emission of each spectrum determined and then the spectrum were transferred to the quasars in the rest frame observer. The step of wavelength to the whole spectrum in the frame of reference in each pixel is limited to 0.5Å and each of the spectra has been normalized. To obtain the normalized spectra at first the mean flux of 21 pixels in the range of 1280Å should be obtained and to divide by each quasar. The spectrum obtained with MATLAB software normalizing the spectrum with the condition that omitting the absorption lines and connecting only the emission lines.

To predict the continuum is sufficient to consider the red region of the spectrum using this part of the spectrum and complete spectrum of quasars of both regions were estimated thus the spectra fitted with weights c_{ij} with PCA method for taking into account the large variety of sample alone was not enough, so an additional suitable parameters must be introduced (see figures 1a & 1b). Then using the continuum to find the best fit parameters a set of free parameters of spectrum and noise (c_{ij}, a_{MF} and b_{MF} - a_{MF} and b_{MF} are free parameters to regulate the mean flux) are considered elementary spectrums are (Lee et al., 2012). However, while the quasar spectra is generally well defined, the red parts of the spectra of Lyα in the spectra in many cases inside the metal absorption lines can be seen in the wavelength range $\lambda_{Rest} \approx 1216 - 1600\text{Å}$. To avoid this characteristic absorption lines feature obtaining from the PCA fitting method (Dall'Aglio et al., 2009).

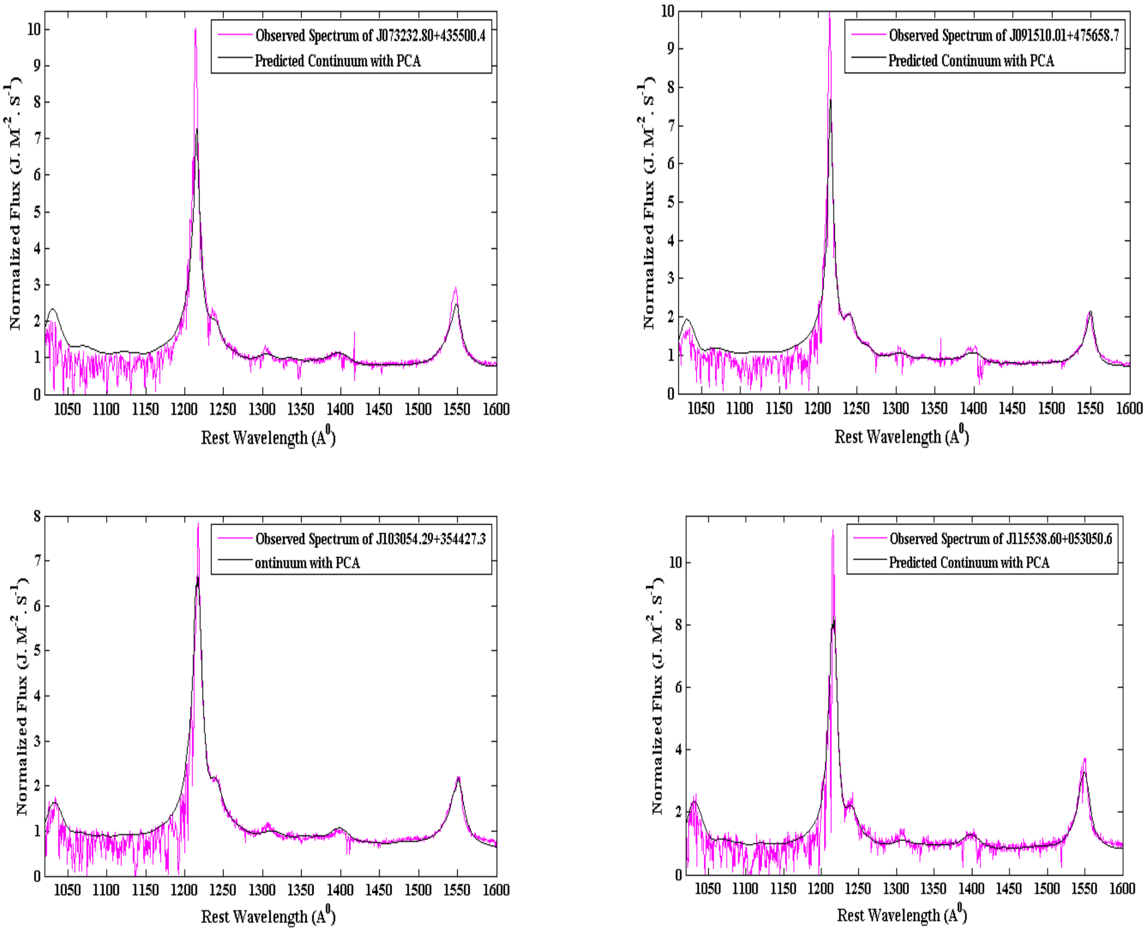

Figure 1b. The good continua of 4 QSOs derived from applying the PCA method described in the text.

In Figures 1a & 1b the continua in PCA method is shown with black lines, we see that the predicted continuum of the PCA method will be observed and work well.

Mean flux regulation

The step in PCA fitting method by continua described in the previous section. However, we have used no information toward the blue parts of the quasar the Lyα emission line due to the absorption from the Lyα forest. Since the absorption redshift at any point is known in the Lyα forest ($z_{Obs} = \lambda_{Obs}/(1216\text{Å} - 1)$), the mean absorption each spectral line can be used to limit the predicted continuum.

In this section the usage of the mean flux evolution of the Lyα forest $\langle F\rangle(z)$ to regulate the amplitude and slope of the predicted PCA continuum were described. In order to regulated the mean flux, using the continuum which obtained by PCA ($C_{PCA}(\lambda)$ fitted to the observed range) $\lambda_{Rest} = 1041 - 1185\text{Å}$ were extracted. Then this divided into bins and the mean flux $\langle F\rangle$ were determined for each bin so $\lambda_{bin} = [1070,\ 1110, 1050\text{Å}]$ are central wavelengths of each bin.

All we need to do is to introduce a quadratic function of a pivot point $\lambda_{Rest} = 1280\text{Å}$, to obtain the mean flux regulated continuum:

$$C_{MF}(\lambda_{Rest}) = C_{PCA}(\lambda_{Rest}) \times (1 + a_{MF}\hat{\lambda}_{Rest} + b_{MF}\hat{\lambda}^2_{Rest}) \qquad (2)$$

Where a_{MF} and b_{MF} are free parameters to fit the regulated range while $\hat{\lambda}_{Reast} = \lambda_{Rest}/1280\text{Å} - 1$ for the regulated and restricted mean flux.

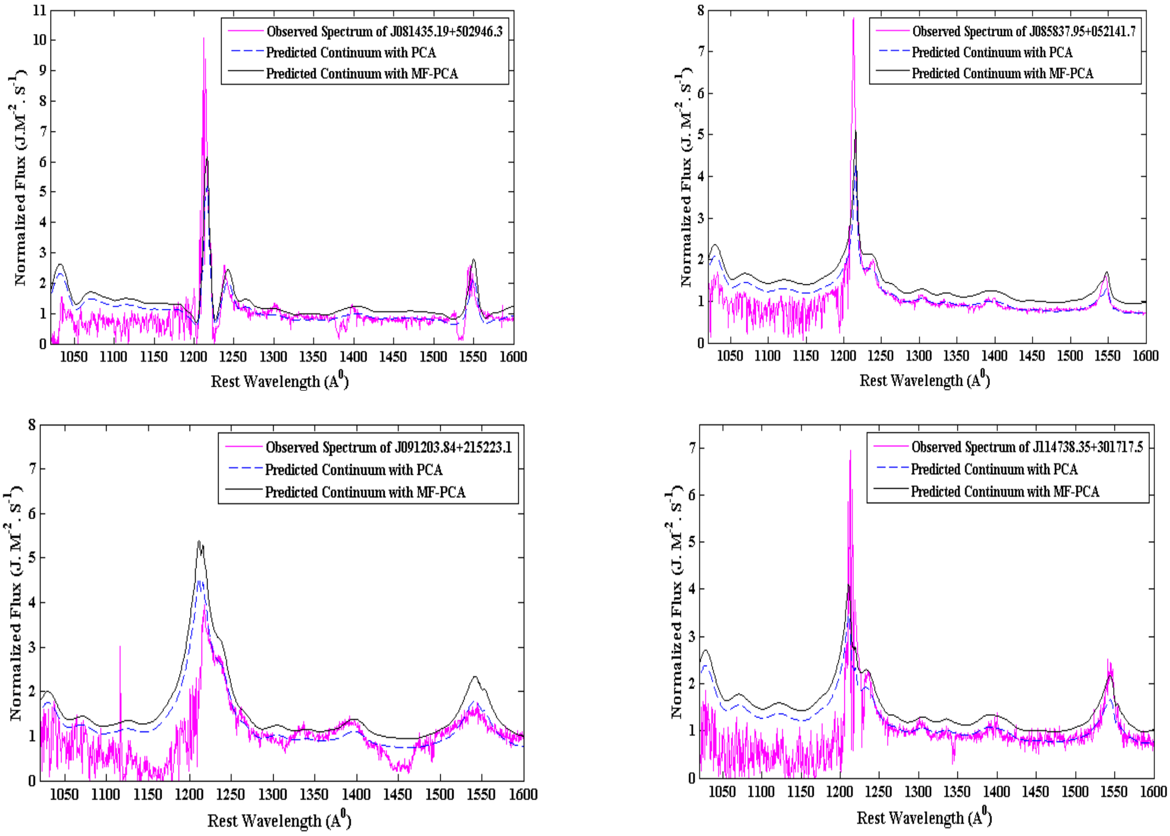

Figure 2a. The bad continua of 4 QSOs derived from applying the MF-PCA method described in the text.

Now the fitting errors (caused by a combination of large scale variance in the Lyα forest) are a few percent. In the next section a mock spectra to quantify the errors in MF-PCA method was discussed (Lee et al., 2012 and Aghaee et al., 2010).

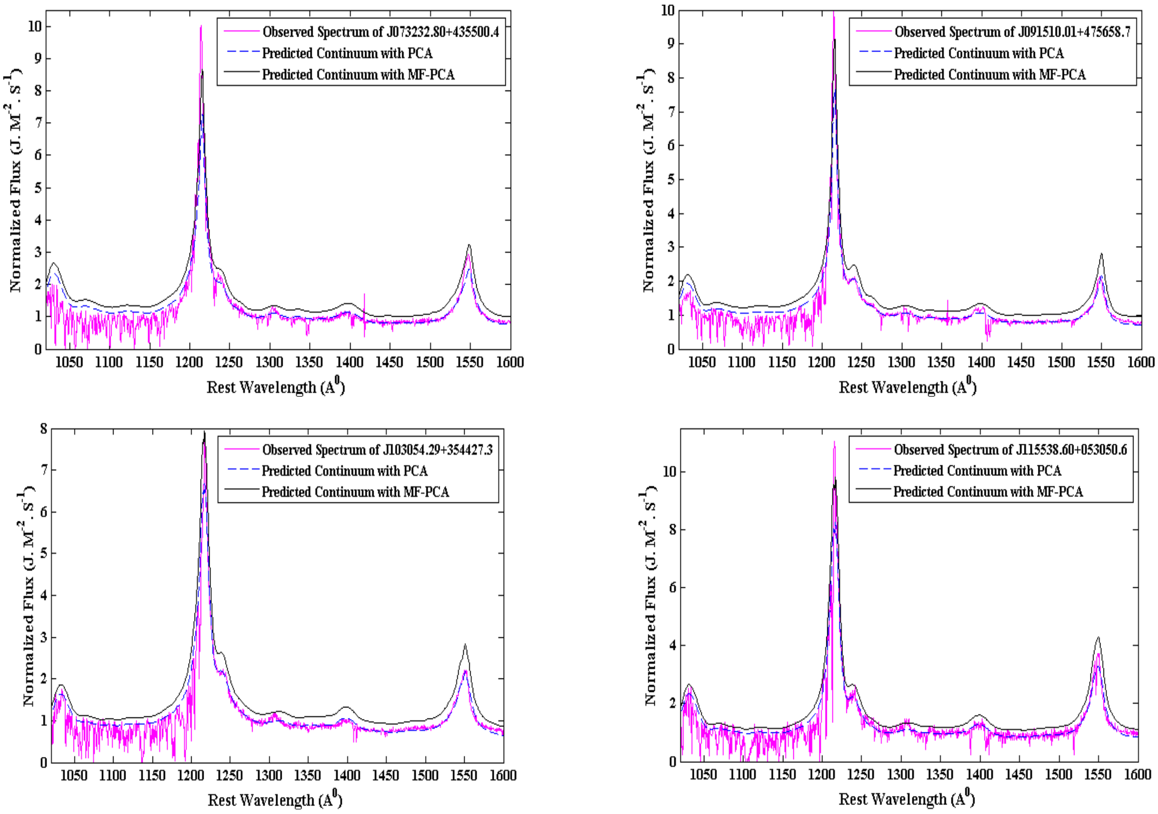

Figure 2b. The good continua of 4 QSOs derived from applying the MF-PCA method described in the text.

Generating artificial spectra

The fitting errors in the MF-PCA method can be tested by the above procedure on noisy mock spectra and comparing the fitted with the correct construction of the spectra is known can be tested. In this section, the actual production of mock spectra and quantitative results of the MF-PCA are explained. The initial step for creating quasar spectra is through specific spectra PCA method and using Gaussian distribution function is c_{ij} weights, as described in article by Suzuki et al., (2006). Note that these results are approximated, because the distribution of the weights cannot be fully Gaussian (while looking for producing real quasar spectra). In particular, the PCA method used to produce the mock spectra and using 10 principal components to increase uncertainty in the fitting, So we do not expect to be able to use (Paris et al., 2011), for fitting the spectra from special spectra (Suzuki et al., 2005) and vice versa. Therefore synthetic quasar spectra fitted through the MF-PCA in the spectral range $\lambda_{Rest} = 1020 - 1600$Å using 10 principal components from the Suzuki et al., (2005) were predicted (Lee et al., 2012 and Aghaee et al., 2010).

Continuum fitting with MF-PCA method

In figures 2a & 2b several examples of the artificial spectra and continuum spectra with fitted MF-PCA method are shown. The first thing to note is that the artificial spectral line looks real.

As described in Section 3, the artificial spectra used as a test for the PCA fit quality on the red side ($\lambda_{Rest} > 1216$Å) of the spectra. According to Figures 2a & 2b, it is clear that the mean flux regulated continuum C_{MF} is modified version of the PCA fit (C_{PCA}). In several cases, the PCA continuum seems nonphysical. These were modified by the mean flux regulated fit by comparing the fitted continue C_{fit} to the true continuum C_{true} which is known by construction in the artificial spectra then the relative error of the continua defined as this way:

$$\delta C(\lambda_{Rest}) \equiv (C_{fit}(\lambda_{Rest})/C_{true}(\lambda_{Rest})) - 1 \quad (3)$$

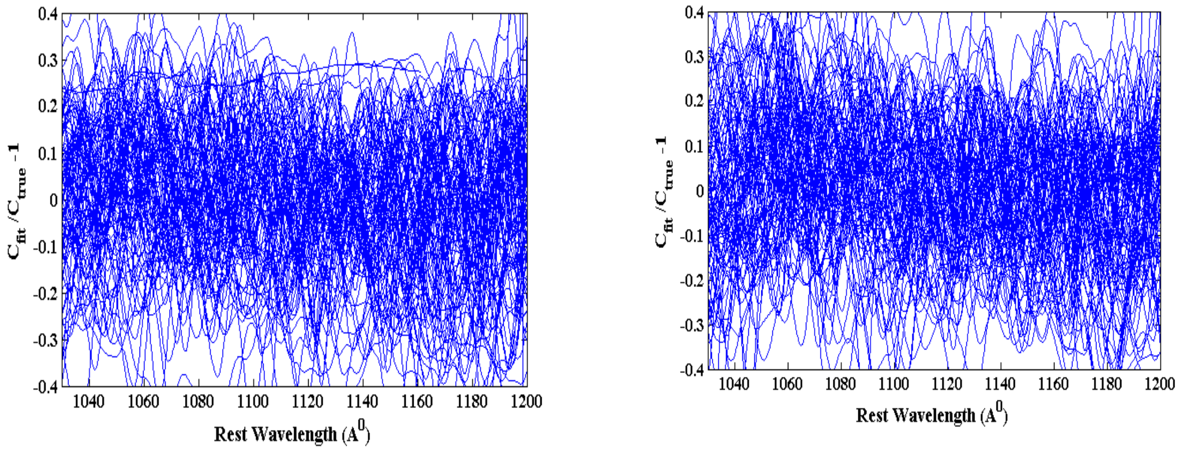

Figure 3. Continuum fitting errors from MF-PCA fitting on mock Lyα forest spectra for initial 200 QSOs.

This can be carried out MF-PCA fits on large numbers of artificial spectra to obtain statistics on the continuum fitting errors as a function of signal-to-noise ratio and quasar redshift (Lee et al., 2012). Figure 3 shows the relative error from fitting of 200 artificial spectra in redshift line. The blue lines are the residuals as a function of wavelength ($\delta C(\lambda_{Rest})$), created into rest frame from a subset of 200 shown. To set a new quality fitted spectrum, the new RMS error of continuum in the wavelength range $\lambda_{Rest} = 1041 - 1185$Å can be defined as:

$$(\delta c)_{rms} = \left[\frac{\int\left(\frac{c_{fit}(\lambda_{Rest})}{c_{true}(\lambda_{Rest})}-1\right)^2 d\lambda_{Rest}}{\int d\lambda_{Rest}}\right]^{\frac{1}{2}} \quad (4)$$

Conclusion

In this paper we matched the precise estimates of Suzuki et al., (2005) and Lee et al., (2012) or Paris et al., (2011) templates. We have estimated the continuum fitting error with in the Lyα forest as a function of quasar redshift and spectral with high signal to noise ratio using mean flux regulated PCA (MF-PCA) a new technique for predicting the Lyα forest continuum. This technique also reduces Fourier power from continuum fitting by a factor of a few in comparison with dividing by a mean continuum. The best continua fit obtained in figures 2a & 2b from 892 quasars. As expected the values of RMS error of 2 to 6% and the average RMS of $2.7 \leq z \leq 4.7$, S/N > 10 and $M_B \leq -28$ is about 4% (See figure 4). So it can be said that at the fixed redshift (z~3) average RMS to with an increase signal to noise ratio decreases at lower redshifts of spectra with low signal to noise ratio, RMS error is relatively high because the regulated mean flux to increase noise levels near the blue end of the SDSS spectra associated with λ_{Obs}~4000Å. This shows that even in the spectra quasars with the signal-to-noise limited the predicted errors in the continua of the PCA method is 3 to 4% error.

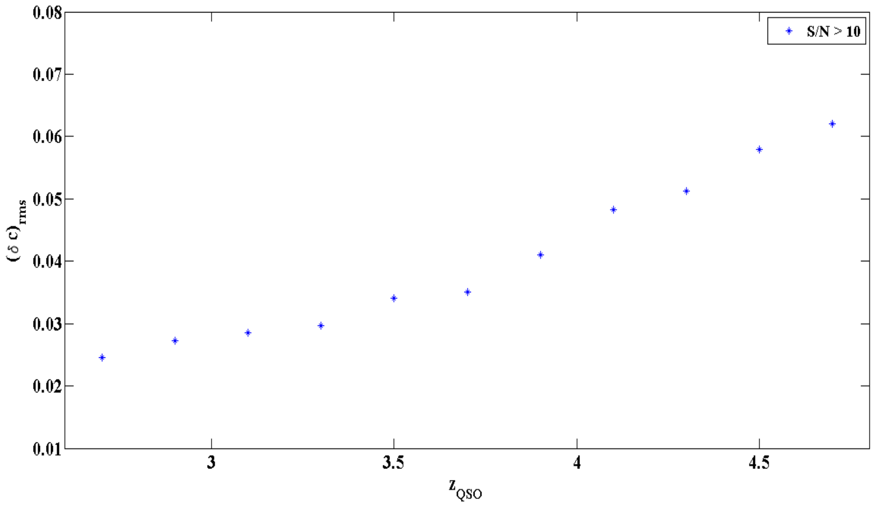

Figure 4. The RMS continuum fitting error in bins of Δz = 0.2 calculated for S/N > 10 and a redshift range 2.7≤ z ≤ 4.7 on the red part of the spectrum (see text).

References

Aghaee, P., Petitjean, S., Stalin, R. Guimaraes and B. Aracil (2006). Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics, 315.

Aghaee, P., Petitjean, R., Srianand, S., Stalin and R. Guimarães (2010). Determination of the Mean HI Absorption of the Intergalactic Medium, Journal of Astrophysics and Astronomy, 31, 59-79.

Becker et al.(2001). Evidence for Reionization at z~6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar, The Astronomical Journal, 122, 2850-2857.

Bernardi et al.(2003). A Feature at z~3.2 in the Evolution of the Lyα Forest Optical Depth, The Astronomical Journal, 125, 32-52.

Dall'Aglio, Wisotzki L. and G. Worseck (2009). The UV background photoionization rate at 2.3<z<4.6 as measured from the Sloan Digital Sky Survey, Astrophysical Journal, 1484, 1-14.

Efstathiou, M., Rowan-Robinson and A. Verma (2002). Submillimetre observations of hyperluminous infrared galaxies, Monthly Notice of the Royal Astronomical Society, 335, 1163-1175.

Francis, C., Hewett, B., Foltz and H. Chaffee (1992). An objective classification scheme for QSO spectra, Astrophysical Journal, Part 1, 398, 476-490.

Hewett, B., Foltz and H. Chaffee (2001). The Large Bright Quasar Survey. VII. The LBQS and FIRST Surveys, the Astronomical Journal, 122, 518-535.

Kirkman et al. (2005). The HI opacity of the intergalactic medium at redshifts 1.6 <z <3.2, Monthly Notices of the Royal Astronomical Society, 360, 1373-1380.

Kim, J., Bolton, M., Viel, G., Haehnelt and R. Carswell (2007). An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties, Monthly Notices of the Royal Astronomical Society, 382, 1657-1674.

Lee, Suzuki, N. and N. Spergel (2012). Mean-flux-regulated Principal Component Analysis Continuum Fitting of Sloan Digital Sky Survey Lyα Forest Spectra, The Astronomical Journal, 143, 51.

Nusser, and M. Haehnelt (2000). The amplitude of mass density fluctuations at z~3.25 from the Lyα forest of Q1422+231, Monthly Notices of the Royal Astronomical Society, 313, 364-376.

Nicolson (1999) Unfolding our Universe, Cambridge University Press, 294.

Press and B. Rybicki (1993). Properties of High-Redshift Lyman-Alpha Clouds. II. Statistical Properties of the Clouds, Astrophysical Journal, 418, 585.

Petitjean and C. Ledoux (1999) Damped Lyman α Systems", Astrophysics and Space Science, 265, 483-486.

Paris et al., (2011). A Principal Component Analysis of quasar UV spectra at Z~3, Astronomy & Astrophysics, 530, 50.

Rauch, W., Sargent and A. Barlow (1998) Glimpses of the Interstellar Medium at Redshift 3.5, eprint arXiv:astro-ph/9812027.

Schneider M. Schmidt and E. Gunn (1991). PC 1247 + 3406 - an optically selected quasar with a redshift of 4.897, Astronomical Journal, 102, 837-840.

Songaila and L. Cowie (1996). Metal enrichment and Ionization Balance in the Lyman Alpha Forest at z=3, Astronomical Journal, 112, 335.

Suzuki, D., Tytler, D., Kirkman, M., O'Meara and D. Lubin (2003). Relative Flux Calibration of Keck HIRES Echelle Spectra, The Publications of the Astronomical Society of the Pacific, 115(811), 1050-1067.

Suzuki D. Tytler D. Kirkman M. John and D. Lubin (2005). Predicting QSO Continua in the Lyα Forest, Astrophysical Journal, 618,592-600.

Strauss (2002). Science Breakthroughs with the Sloan Digital Sky Survey: Results and Opportunities, Survey and Other Telescope Technologies and Discoveries. Edited by Tyson, J. Anthony; Wolff, Sidney. Proceedings of the SPIE, 4836, 1-9.

Seljak and R. Cen (2003). Precision cosmology from the Lyman α forest: power spectrum and bispectrum, Monthly Notices of the Royal Astronomical Society, 344, 776-788.

Whitney W. Porcas and A. Preston (1983). Detection of a compact radio source near the center of a gravitational lens - Quasar image or galactic core, Science, 219, 54-56.

White, J., Helfand, H., Becker, D., Gregg, M., Postman, R., Lauer and W. Oegerle (2003). An I-Band-selected Sample of Radio-emitting Quasars: Evidence for a Large Population of Red Quasars, The Astronomical Journal, 126, 706-722.